 Human Molecular Genetics, 1999, Vol. 8, No. 5 775–782

ARTICLE

The RNA-binding properties of SMN: deletion analysis of the zebrafish orthologue defines domains conserved in evolution

Solange Bertrandy, Philippe Burlet, Olivier Clermont, Céline Huber, Christian Fondrat[1], Danielle Thierry-Mieg[2], Arnold Munnich and Suzie Lefebvre*

Unité de Recherches sur les Handicaps Génétiques de l'Enfant, INSERM U393, IFREM, Institut Necker, Hôpital Enfants Malades, 149 rue de Sèvres, 75743 Paris cédex 15, France, [1]Citi2, 75006 Paris, France and [2]CRBM-CNRS, 34293 Montpellier cédex 5, France

Received January 18, 1999; Revised and Accepted February 24, 1999

DDBJ/EMBL/GenBank accession nos Y17255 and Y17256

Spinal muscular atrophy (SMA) is a common autosomal recessive disorder that results in the degeneration of spinal motor neurons. SMA is caused by alterations of the survival motor neuron (*SMN*) gene which encodes a novel protein of hitherto unclear function. The SMN protein associates with ribonucleoprotein particles involved in RNA processing and exhibits an RNA-binding capacity. We have isolated the zebrafish *Danio rerio* and nematode *Caenorhabditis elegans* orthologues and have found that the RNA-binding capacity is conserved across species. Purified recombinant SMN proteins from both species showed selectivity to poly(G) homopolymer RNA *in vitro*, similar to that of the human protein. Studying deletions of the zebrafish SMN protein, we defined an RNA-binding element in exon 2a, which is highly conserved across species, and revealed that its binding activity is modulated by protein domains encoded by exon 2b and exon 3. Finally, the deleted recombinant zebrafish protein mimicking an SMA frameshift mutation showed a dramatic change *in vitro* in the formation of the RNA–protein complexes. These observations indicate that the RNA-binding capacity of SMN is an evolutionarily conserved function and further support the view that defects in RNA metabolism most likely account for the pathogenesis of SMA.

INTRODUCTION

Spinal muscular atrophy (SMA) is a common autosomal recessive neuromuscular disease characterized by degeneration of motor neurons of the spinal cord, leading to progressive muscular atrophy. Three forms of childhood SMA are recognized on the basis of age of onset and clinical course (1). The SMA phenotype results from alterations of the survival motor neuron (*SMN*) gene (2). Indeed, the *SMN* gene is absent in 98.6% of SMA patients, and cases retaining the gene carry intragenic mutations (reviewed in refs 3,4). The *SMN* gene and its centromeric copy (*SMNc*) are located in an unstable duplicated region of chromosome 5q13 (2,5). No such duplication has been detected in mouse, suggesting that duplication of the human *SMN* gene results from a recent event (6,7). The *SMN* and *SMNc* genes differ by five nucleotide substitutions and encode a 294 amino acid protein of hitherto unclear function (2,8). A tight correlation between the level of SMNc protein and the severity of disease has been observed in tissues derived from SMA patients (9–11).

The SMN protein is present both in the cytoplasm and in a novel nuclear structure, called *gem* (12), involved in RNA processing in particular during spliceosomal snRNP biogenesis and pre-mRNA splicing (13–15). SMN has been shown to contain two highly conserved domains (16) found to bind to the SMN-interacting protein 1 (SIP1) and Sm proteins, respectively (13; Fig. 1). The Sm-interacting domain was shown to overlap the self-oligomerization domain (17). A domain of *SMN* exons 2a–2b has also been shown to mediate nucleic acid-binding activity (18). Thus, a better knowledge of conserved domains among *SMN* orthologue genes should provide further insights into functional and structural requirements of SMN.

We have isolated the zebrafish *Danio rerio* and nematode *Caenorhabditis elegans SMN* orthologue genes. Computer predictions identified in SMN proteins a sequence similar to the ribonucleoprotein consensus sequence (RNP1), which is a highly conserved octapeptide core usually found in RNA- (19) and DNA-binding proteins (20). Recombinant SMN proteins from both species were produced and exhibited a conserved RNA-

*To whom correspondence should be addressed. Tel: +33 1 44 49 51 63; Fax: +33 1 47 34 85 14; Email: lefebvre@necker.fr

Figure 1. Comparison of the SMN orthologue proteins. Alignment of the predicted amino acid sequences from human (2), mouse (6), zebrafish and *C.elegans SMN* genes. Conserved residues are shaded. The conserved regions corresponding to amino acids 13–44 and 240–267 of human SMN protein are protein-binding sites for SIP1 and Sm proteins, respectively (13,15).

binding capacity. Deletions of the zebrafish SMN protein allowed us to identify domains modulating the binding properties of the SMN RNA-binding determinant. This high conservation of the SMN RNA-binding capacity during evolution indicates an essential role for the protein in RNA metabolism. Further biochemical and genetic investigations hopefully will help to identify sequence-specific RNA ligands and elucidate essential SMN functions in the pathogenesis of SMA.

RESULTS

Detection and identification of *SMN* orthologue genes

A computer search using the human SMN protein detected a genomic sequence from *C.elegans* in the GenBank database (cosmid C41G7). The alignment of human, mouse and *C.elegans* sequences revealed two highly conserved domains at the N- and C-termini, respectively (16; Fig. 1). Degenerate primers were designed to bind these domains and used for PCR amplification of the *SMN* cDNAs from worm to vertebrates, indicating that they should allow isolation of the *SMN* orthologues (Fig. 2).

Despite identification of a homologous *SMN* sequence in the *C.elegans* genome, there was no previously reported cDNA clone (21). The *C.elegans* cDNA was cloned using specific primers, and sequence analyses showed a homology with different mammalian *SMN* sequences (Table 1, Fig. 1). The full-length cDNA (699 bp) contained a 591 nucleotide open reading frame (ORF) encoding a 197 amino acid protein shorter than the human counterpart (294 amino acids). The predicted *C.elegans* SMN protein shows 36% homology with the human protein. *Trans*-splicing is responsible for addition of a spliced leader sequence onto the 5′ end of many nematode mRNAs (22). Sequence analyses further revealed that the *C.elegans SMN* mRNA is not *trans*-spliced. The cDNA has a 3′-untranslated region (3′-UTR) sequence of 96 bp which contains a polyadenylation signal (AATAAA) followed by a poly(A) tail.

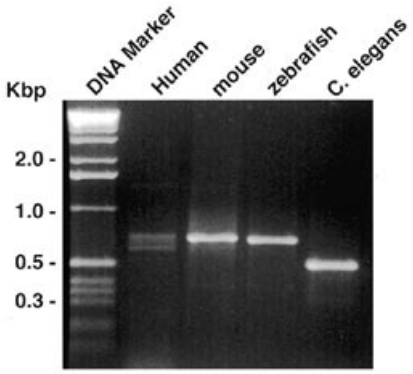

Figure 2. High evolutionary conservation of the *SMN* gene. RT–PCR analyses of *SMN* RNA transcripts from different species. Electrophoresis of RT–PCR amplification products on a 2% agarose gel revealed a major DNA band of ~800 or 500 bp, corresponding to human, mouse and zebrafish, or *C.elegans*, respectively. Additional bands observed for human *SMN* transcripts corresponded to predicted alternatively spliced forms (2,23).

Table 1. Quantitative analyses of *C.elegans* and zebrafish *D.rerio* orthologue SMN proteins

	Percentage identity (and similarity)			
	Human	Mouse	Zebrafish	*C.elegans*
Human	100	83 (88)	52 (66)	36 (56)
Mouse	–	100	52 (66)	38 (58)
Zebrafish	–	–	100	37 (58)
C.elegans	–	–	–	100

The human (2) and mouse (6) SMN proteins have been reported previously.

To isolate and characterize the zebrafish *SMN* cDNA, an RT–PCR amplification product obtained with the degenerate oligonucleotide primers was sequenced (Fig. 2). Then, the 5′ and 3′ ends of the full-length cDNA molecule were obtained by rapid amplification of cDNA ends using specific primers. The cDNA molecule has a short 5′-UTR (71 bp) and a longer 3′-UTR (100 bp) that contains a polyadenylation signal (AATAAA) 25 bp from the poly(A) tail. The full-length zebrafish cDNA (1016 bp) contains a 843 nucleotide ORF encoding a 281 amino acid protein slightly shorter than the human counterpart (294 amino acids). A BLAST search of the GenBank database detects an expressed sequence tag (EST) (accession no. AA494875) from the zebrafish genome project that corresponds to the 5′ end of our clone and extends 25 bp further upstream. This EST is likely to encompass the full-length cDNA as indicated by the sequence of the 3′ end (accession no. AA494767). Recently, a zebrafish cDNA clone (accession no. AF083557) has also been identified and reveals an ORF (285 amino acids) that has 95% identity with the ORF described herein. Sequence comparison showed that the zebrafish amino acid sequence predicted from our clone has a 52% homology with the human and 37% with the *C.elegans* SMN protein (Table 1, Fig. 1).

Characterization of the *SMN* orthologue genes

The zebrafish *SMN* gene is composed of nine exons encompassing ~8.5 kb, and its structure is similar to that of the human *SMN* gene (6; Fig. 3a). It is interesting to note that the nematode *C.elegans SMN* gene is converted into its zebrafish orthologue by addition of coding exons 1, 5, 7 and part of exons 2b and 4. In addition, we identified a (CA) microsatellite DNA marker (*smn5i*) in intron 5 which should help in mapping the zebrafish *SMN* gene. RT–PCR amplification of adult zebrafish poly(A)$^+$ RNA failed to reveal alternatively spliced transcripts such as those detected in human tissues (2,23). Similar results have been reported for the mouse *SMN* gene (6,7). FASTA and BLAST searches failed to reveal any significant homology with proteins other than the SMN orthologue sequences. There are two highly conserved regions in the N- and C-terminal ends of the SMN protein that are involved in protein interactions with SIP1 and the Sm proteins, respectively (13; Fig. 1).

Comparative analyses with protein patterns from PRODOM and PROSITE databases showed a weak similarity with a domain of the tudor protein. This homology has been reported previously for the human and *C.elegans* SMN proteins (24). The best score was observed with the SMN PRODOM domain defined by the first 194 amino acids of the human protein. A comparative search in the PROSITE database using one mismatch at any amino acid position identified a sequence similar to the RNP1 motif found in the DNA-binding cold shock domain (20) and in the RNA-binding domain (RBD) of many RNA-binding proteins (19; Fig. 3b). The RBD is usually composed of two highly conserved motifs: a hydrophobic segment of six residues (the RNP2 motif) and an octapeptide motif (the RNP1 motif). The RNP1 motif is characterized by a β-sheet structure and, despite limited similarity over the eight amino acid residues, the conserved β-sheet core was apparent in the RNP1-like motif of the orthologue SMN proteins, but the SMN proteins lacked the canonical RNP2 motif. This observation suggests that an RNA-binding domain is present in the SMN protein.

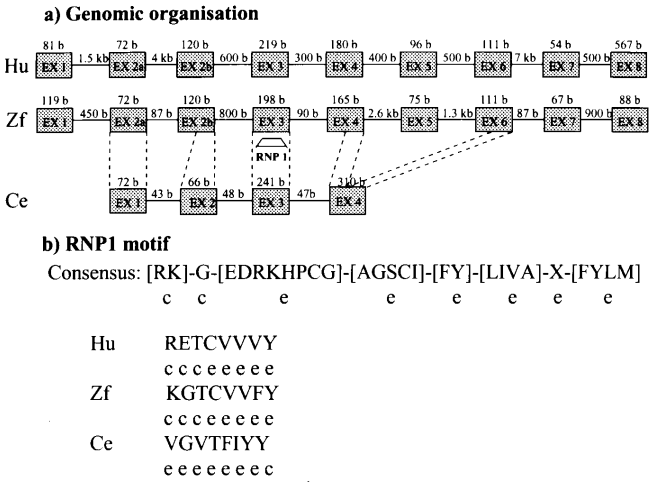

Figure 3. Structural analyses of *SMN* genes and proteins. (**a**) Comparison of the genomic structures of the human, zebrafish and *C.elegans SMN* genes. The zebrafish *SMN* orthologue gene consists of nine exons, compared with four exons in the *C.elegans SMN* orthologue gene. (**b**) The RNP1 consensus sequence is indicated on the top of the alignment (Prosite, PD0C00030). The RNP1-like motif is depicted with the predicted position of secondary structural element β-sheets (e) and coiled-coils (c) in the human (Hu), zebrafish (Zf) and *C.elegans* (Ce) SMN proteins.

An RNA-binding determinant lies at the N-terminal end of the SMN protein

To characterize the RBD of the zebrafish SMN protein, we investigated whether this protein could bind an RNA transcript from the polylinker of pBluescript. A series of SMN deletion constructs was designed and the corresponding recombinant proteins were tested for their ability to bind RNA under our experimental conditions (Fig. 4A). The recombinant proteins were fused to both an N-terminal His6 and S-peptide tag that allowed their immunoblot detection (Fig. 4B). The northwestern analyses showed that the deletion of the first 105 N-terminal amino acids of the full-length protein (ZF-FL) completely abolished the RNA-binding activity of the recombinant protein (construct ZFSMN4; Fig. 4C, Table 2). The ZFSMN1 construct, which contains the first 76 N-terminal amino acids, interacted with RNA very efficiently *in vitro*, indicating that the putative RNP1 motif was not essential under the present binding conditions. Interestingly, when studying the binding activity of a deletion mimicking an SMA frameshift mutation (25), we observed that deletion of the last 138 C-terminal amino acids (construct ZFSMN3) also completely abolished RNA binding, whereas a construct deleted of the last 177 C-terminal amino acids showed a significant RNA-binding activity (construct ZFSMN2).

A series of binding conditions was tested to characterize the interaction of the isolated SMN RNA-binding determinant (construct ZFSMN1). Studying the effect of ionic strength on the stability of the RNA–protein interaction (Fig. 5A), we found that the SMN RNA-binding determinant exhibited a binding activity at 100 and 250 mM NaCl. The RNA association was still detectable at 0.5 and 1 M NaCl. Then, competition experiments were carried out using *Escherichia coli* tRNAs as competitors (Fig. 5B). The presence of 2, 10 and 100 µg/ml of tRNAs did not completely abolish the RNA binding. At a higher concentration

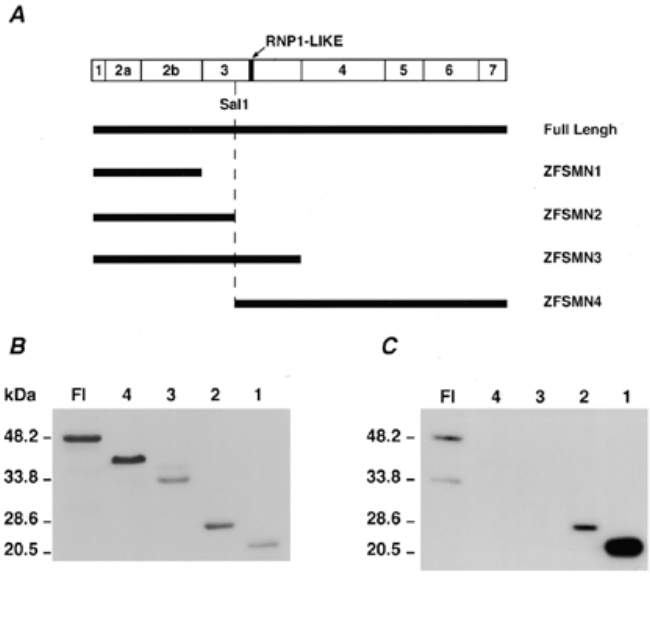

Figure 4. RNA-binding activity of deletions of the zebrafish SMN protein. (**A**) Schematic representation of the full-length and deleted zebrafish SMN proteins. The RNP1-like motif and the restriction enzyme site for *Sal*I are indicated. (**B**) Recombinant proteins were analysed by immunoblotting using an anti-tag detection system to estimate the amount of input protein. Fl, full-length. (**C**) Northwestern blotting was carried out with the same membrane incubated with radiolabelled RNA transcript from the polylinker of pBluescript. The SMN RNA-binding determinant lies in the N-terminal end of the protein, outside the RNP1-like motif. Fl, full-length.

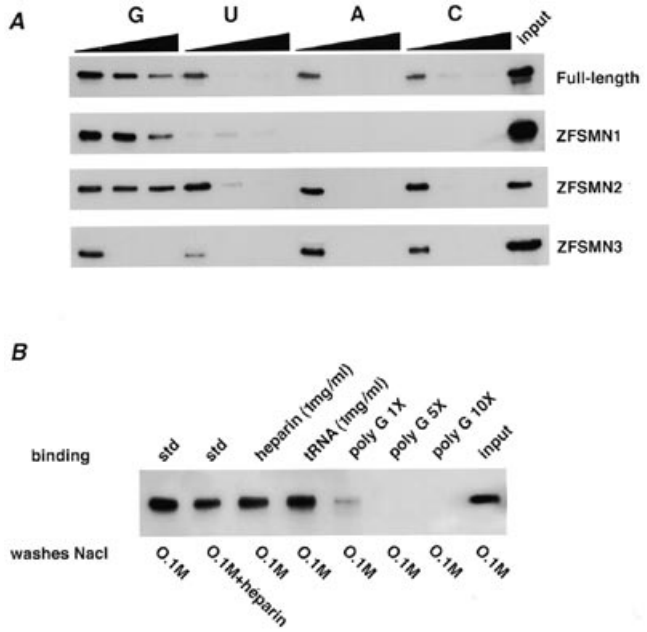

Figure 6. Zebrafish SMN protein selectively binds to ribonucleotide homopolymers. (**A**) Purified recombinant proteins were incubated with each homopolymer and washed under increasing NaCl concentrations (0.1, 0.25 and 0.5 M). The proteins retained by the immobilized homopolymer RNAs were analysed by immunoblotting. (**B**) Characterization of the poly(G) RNA interaction with the SMN RNA-binding determinant (ZFSMN1). Purified recombinant protein was mixed with immobilized poly(G) under various conditions as indicated. The protein retained was washed under the conditions described for each lane.

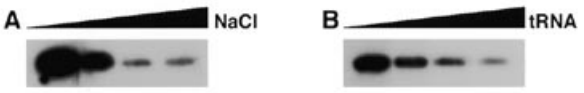

Figure 5. Characterization of the SMN RNA-binding determinant. Northwestern experiments were performed using purified recombinant ZFSMN1 protein and RNA transcribed from the polylinker of pBluescript. The effects of (**A**) increasing salt concentrations (0.1, 0.25, 0.5 and 1 M) and (**B**) increasing concentrations of *E.coli* tRNA (2, 10, 100 and 1000 µg) used in binding assays.

(1 mg/ml tRNA), the RNA probe was largely removed but some interaction was still detectable. These results indicated that the SMN protein binds RNA in a sequence-independent manner.

The orthologue SMN proteins have RNA-binding preference

To characterize further the RNA-binding properties of the zebrafish SMN protein, we performed binding assays with ribonucleotide homopolymers conjugated to agarose beads. For these experiments, the recombinant proteins were overexpressed in *E.coli*, purified under native conditions and used for binding assays. The proteins bound to the homopolymer–beads were washed in a buffer with increasing NaCl concentrations (0.1, 0.25 and 0.5 M) to assess the binding preference (Fig. 6, Table 2). These experiments showed that the SMN protein binds more strongly to poly(G) than to poly(U), poly(A) or poly(C) (Fig. 6A). The binding of SMN to the poly(U), poly(A) and poly(C) homopolymer ligands was abolished by increasing the ionic strength, accounting for the role of ionic contacts in RNA–protein interactions. The resistance of the protein–poly(G) interaction to ionic strength suggested that ionic contacts may not contribute significantly to this interaction, but rather that hydrophobic interactions are involved in the formation of RNA–protein complexes. Thus, the binding capacity of the ZFSMN1 protein revealed that the RNA-binding determinant does not have the binding properties of the full-length SMN protein (Fig. 6A). Deleting exon 2b did not abolish the binding to poly(G), but affected its resistance to ionic strength (data not shown). The binding properties of SMN are reflected reliably by its derivative ZFSMN2, suggesting a cooperative involvement of the region C-terminal to the RNA-binding determinant. Interestingly, ZFSMN3 protein did not bind preferentially to poly(G), suggesting that it has lost the ability to distinguish between the four ribonucleotide homopolymers. These observations supported the view that adjacent protein domains within SMN exon 3 might modulate the RNA-binding activity.

To characterize further the stability of the SMN–poly(G) complex, interaction with the SMN RNA-binding determinant (ZFSMN1 protein) was carried out under various experimental conditions (Fig. 6B). The ZFSMN1 protein retained its binding capacity in the presence of either 1 mg/ml of *E.coli* tRNA or heparin. The binding specificity was assessed by performing the binding reaction in the presence of free poly(G) competitor. An equimolar amount of poly(G) competitor markedly reduced the amount of protein bound to poly(G) conjugated to agarose beads,

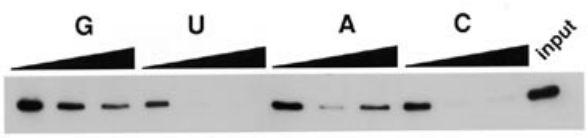

Figure 7. *Caenorhabditis elegans* SMN protein selectively binds to ribonucleotide homopolymers. The purified recombinant protein was incubated with each homopolymer and washed under increasing NaCl concentrations. The protein retained by the immobilized homopolymer RNAs was revealed by immunoblotting analyses.

and a 10-fold molar excess completely hampered the protein binding to the poly(G)–beads.

These results raised the interesting possibility that the RNA-binding capacity of SMN protein is conserved during evolution. To test this hypothesis, we performed binding assays with ribonucleotide homopolymers using the recombinant *C.elegans* SMN (CeSMN) protein purified after overexpression in *E.coli*. The protein retained on the homopolymer–beads was washed in buffer with increasing NaCl concentrations to define the binding preference (Fig. 7). The CeSMN protein showed the same preference for poly(G) as the ZFSMN protein. In addition, the CeSMN protein appeared to be a polypurine-binding protein because it also tightly bound poly(A) homopolymer RNA, suggesting a direct role of SMN in the maturation of mRNA.

Table 2. Summary of the RNA-binding properties of deleted zebrafish SMN proteins

Proteins[a]	Northwestern	Poly(A)	Poly(C)	Poly(G)	Poly(U)
ZF-FL	+	+	+	+++	+
ZFSMN1	+++	–	–	+++	+/–
ZFSMN2	++	+	+	+++	+
ZFSMN3	–	+	+	+	+
ZFSMN4	–	ND	ND	ND	ND

Purified zebrafish SMN recombinant proteins corresponded to full-length (ZF-FL), the first 76 N-terminal amino acids (ZFSMN1), the first 104 N-terminal amino acids (ZFSMN2), the first 143 N-terminal amino acids (ZFSMN3) and the last 176 C-terminal amino acids (ZFSMN4).
ND, not done.

DISCUSSION

The identification of the *SMN* gene has represented an important step towards the goal of elucidating the molecular basis of SMA (2). The SMN protein is markedly deficient in tissues from SMA fetuses and patients (9–11), and early embryonic lethality is observed in SMN knock-out mice (26). Moreover, biochemical studies have shown that the SMN complex is involved in spliceosomal snRNP biogenesis (13,14) and pre-mRNA splicing (15). Yet, the specific function of SMN remains unclear and the identification of evolutionarily conserved domains and protein motifs should, therefore, help to devise experiments aimed at understanding its function(s). Previous studies have shown that SMN interacts with both RNA-binding proteins and RNA (12–15). The two most conserved domains identified in the N- and C-terminal ends of SMN have been found to bind SIP1 and Sm proteins, respectively (13). The Sm-interacting domain was shown to overlap the self-oligomerization domain (17). The RNA-binding capacity of SMN in humans is accounted for by exons 2a–2b (18). However, the SMN–RNA interactions across species remain to be established.

Here, we have isolated the *C.elegans* and zebrafish *SMN* orthologues and have shown that the corresponding recombinant proteins interact with RNA directly. Deletion studies mapped the RNA-binding domain in the N-terminal end of SMN. The zebrafish *SMN* exon 2a has 100% homology to human exon 2a, while exon 2b is only 38% homologous to the human counterpart. Sequence comparison with *C.elegans* SMN has allowed the delineation of a highly conserved region of exon 2a which overlaps with the SIP1-interacting domain in human (13; Fig. 8A). Screening the *C.elegans* genome sequence for SIP1 failed to reveal any homologous proteins (27), favouring the view that at least part of the highly conserved domain in the N-terminal end of SMN may reflect its conserved RNA-binding capacity. It is interesting to note that this conserved domain is predicted to form an amphipathic α-helix (Fig. 8B), suggesting that the opposite sides of the helix may be involved in different interactions. The resistance of the SMN–RNA interaction to ionic strength reported here also suggests that hydrophobic contacts may contribute to the formation of the RNA–protein complexes. Further investigations should help to determine the amino acid residues directly interacting with nucleic acids.

The purified recombinant zebrafish SMN protein exhibited RNA homopolymer preference for poly(G) similar to that previously reported for human SMN produced by *in vitro* translation (18). In the previously reported study, however, an interference with the translation system could not be totally ruled out. Here, the RNA-binding preference is shown to be an intrinsic property of SMN, irrespective of other proteins. Our deletion analyses of the zebrafish SMN favour the view that its RNA-binding activity may be modulated by domains located C-terminal to exon 2a and termed auxiliary domains. The exon 2b domain increased the RNA-binding affinity, suggesting that the charged amino acids may form an ionic interaction with RNA phosphates to stabilize the RNA–protein complex. The candidate positive auxiliary domain formed by the first 24 amino acids of exon 3 might interact with RNA once the SMN RNA-binding determinant has bound the RNA ligand. This auxiliary domain is predicted to include two short β-sheets containing aromatic amino acid residues. One can speculate that they are candidate structures for cooperative RNA interactions. This region is also highly conserved in a novel 30 kDa splicing factor (28,29), indicating that it may serve a general function in RNA-binding proteins. On the other hand, an SMN deletion construct (ZFSMN3) mimicking an SMA frameshift mutation (25) altered the formation of the SMN–RNA complexes, indicating that RNA binding might also be modulated by a negative auxiliary domain located in exon 3. Computer analyses of the region detected a motif similar to the RNP1, which is a highly conserved octapeptide core usually found in RNA-binding proteins (19). Under our experimental conditions, the RNP1-like motif is not essential for RNA binding, but is a good candidate sequence to modulate the formation of SMN–RNA complexes. Further experiments will be necessary to characterize the involvement of auxiliary domains in the RNA-binding properties of SMN when the sequence-specific RNA ligands are available. Similar results

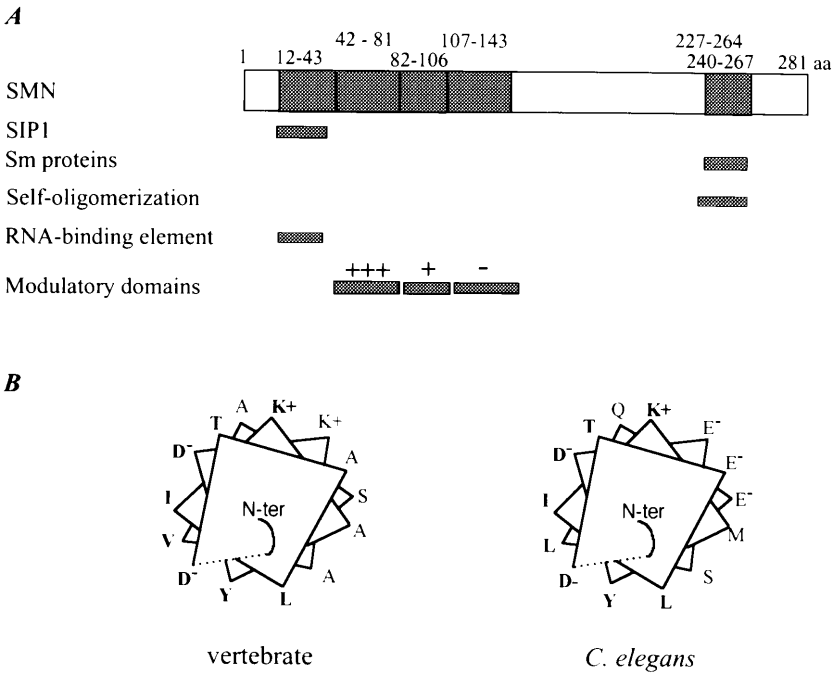

Figure 8. Schematic representation of the SMN-interacting sites. (**A**) Domain structure of the SMN protein. Domains corresponding to amino acids 13–44 and 240–267 bind to SIP1 and Sm proteins, respectively (13,14). SMN exon 6 (amino acids 242–278) directly mediated self-oligomerization (17). The domain comprising amino acids 22–32 accounted for the RNA-binding determinant as reported here for SMN. Candidate auxiliary domains positively (+) and negatively (–) modulating the RNA-binding capacity of SMN are located at amino acids 51–91, 77–105 and 106–143, respectively. (**B**) The secondary structure of the highly conserved N-terminal domain. A helical projection is shown, with the hydrophobic and charged amino acids clustered on opposite sides of the predicted α-helix.

have been reported for other RNA-binding proteins in which the RBD is distinct from the domains that provide affinity and specificity (30). Furthermore, screening for the Sm proteins found highly similar sequences in the *C.elegans* genome, suggesting that the highly conserved domain in the C-terminal end of the *C.elegans* SMN should interact with Sm proteins as reported for the human protein (13). These findings indicated that a deletion of the C-terminal end of SMN, which binds Sm proteins, can alter the RNA-binding capacity of SMN. Therefore, the two binding properties of SMN might be closely related and serve in the assembly of ribonucleoprotein complexes.

These findings, together with the role of SMN in RNA metabolism (12–15) and the reduced RNA-binding activity of SMN mutant proteins (18), give strong support to the view that disruption of RNA processing may account for the pathogenesis of SMA (31). The present study also suggests that purification and refolding of the recombinant SMN protein did not impair its RNA-binding activity. The selective RNA-binding activity of the purified recombinant SMN proteins should allow the isolation and characterization of sequence-specific RNA ligands and help in identifying putative aberrant RNA maturation in SMA. An enormous amount of information is now available on *C.elegans*, which makes this organism one of the best biological models to study gene functions. Further investigations will be required to understand SMN functions in *C.elegans* and should contribute to the elucidation of the role of SMN in cell survival.

MATERIALS AND METHODS

Isolation and cloning of *SMN* orthologue cDNAs

Zebrafish poly(A)$^+$ RNA was isolated from total RNA using the QuickPrep mRNA purification kit (Pharmacia, Uppsala, Sweden). Reverse transcription of either poly(A)$^+$ RNA or total RNA was carried out at 42°C using avian myeloblastosis virus reverse transcriptase (Promega, Madison, WI) and random hexamers (Pharmacia). For PCR amplification of single-stranded cDNA, we used forward primer nex2c (5′-NTNTGGGAYGAY-ACNGCNYT-3′) and reverse primer nex6el (5′-RTGRTANC-CNSWCATRTACCA-3′), denaturation at 94°C for 1 min, followed by 30 cycles of 1 min at 94°C, 1 min at 55°C, 1 min at 72°C and a final extension at 72°C for 4 min. Amplification of the 5′ and 3′ ends of zebrafish *SMN* cDNA was performed using a Marathon cDNA amplification kit (Clontech, Palo Alto, CA) according to the manufacturer's instructions. Briefly, oligo(dT)-based reverse transcription was carried out on poly(A)$^+$ RNA and used as a template for double-stranded cDNA synthesis. The double-stranded cDNA was linked to adaptors which allow 5′ and 3′ end amplification by a semi-long range PCR (Boehringer Mannheim, Meylan, France) using adaptor-specific primers (Clontech) and zebrafish *SMN*-specific primers. The 5′ end was amplified using forward AP2 primer and reverse primer zf10 (5′-GCATTCTTGAATGATGCAACAGC-3′). For

the 3′ end, we used forward primer zf3 (5′-CAGAGTCCTCCA-CAGAAGAGAGT-3′) and reverse AP2 primer. For PCR amplification of the *C.elegans* cDNA, we used reverse-transcribed total RNA and primers derived from the genomic sequence given in the GenBank database (accession no. Z81048). Amplification of the 5′ and 3′ ends of *SMN* cDNA was carried out using a nematode cDNA library (Stratagene, La Jolla, CA) constructed in the Uni-ZAP XR vector. The 5′ end was amplified using forward T7 primer and reverse primer 3Ce1 (5′-TCCAAACATAATCG-CAATGGCC-3′). For the 3′ end, we used forward primer 5Ce2 (5′-TGTACAACTGCTTGACCTCC-3′) and reverse T3 primer. The coding sequence of both nematode and zebrafish *SMN* cDNA molecules was amplified by semi-long range PCR (Boehringer Mannheim) and cloned into PCRscript vector (Stratagene) using standard procedures.

Gene structure

Gene structure analyses were carried out by long-range PCR amplification using exonic primers (Boehringer Mannheim). We identified a (CA) microsatellite marker (*smn5i*) in intron 5 of the zebrafish *SMN* orthologue gene that is detected by PCR amplification using the forward primer 5′-TGTTCCTGTGATCTC-CAATCT-3′ and reverse primer 5′-ATGCACATAACATGCTG-GCG-3′.

DNA sequencing and computer analyses

DNA sequencing of PCR products and double-stranded plasmid DNAs was performed on both strands using the ABI Prism BigDye terminator cycle sequencing ready reaction kit (Perkin-Elmer/ABI, Foster City, CA) and analysed on an Applied Biosystems 373A DNA automated sequencer. The resulting sequences were assembled into a contig and the amino acid ORF identified was aligned with the different SMN orthologue sequences using the CLUSTAL-W program (32). BLAST and FASTA searches were performed against the non-redundant protein databases. Protein patterns were searched for in the PROSITE (33) and PRODOM databases (34) using the RAMdb program (35). The DNA and amino acid sequences of *C.elegans* and zebrafish *SMN* have been submitted to the EMBL database and have been assigned the accession nos Y17255 and Y17256, respectively.

Production of recombinant proteins

DNA fragments obtained by restriction enzyme digestion or by PCR amplification were cloned into vector pET30 (Novagen, Madison, WI) for protein expression in *E.coli* BL21(DE3). Briefly, the recombinant proteins were produced following the addition of isopropyl-β-D-thiogalactopyranoside (1 mM) to bacteria grown in Luria broth at 37°C until the A_{600} reached 0.6–0.8. After a 3 h induction, the bacteria were harvested and lysed in a buffer containing 6 M guanidinium hydrochloride, 0.1 M sodium phosphate, 0.5 M NaCl, 0.01 M β-mercaptoethanol and 0.01 M Tris–HCl, pH 8.0. The lysate was centrifuged and loaded onto an Ni-NTA column (Qiagen, Chatsworth, CA). The proteins were purified under either denaturating (urea) or native conditions (imidazole) according to the manufacturer's instructions.

Preparation of RNA probes

The Bluescript plasmid was used as template for synthesis of a 240 nucleotide RNA transcript. One microgram of *Pvu*II-digested plasmid DNA was transcribed and *in vitro* ^{32}P-labelled with T3 polymerase as recommended by the manufacturer (Boehringer Mannheim). After addition of RNase-free DNase RQ1 (Promega) for 30 min at 37°C, the radiolabelled transcript was fractionated on a chromaspin 10 column (Clontech) and stored at –20°C. Prior to use, the RNA probe was heated for 10 min at 90°C and cooled down to room temperature for 20 min.

Northwestern blot analyses

The recombinant proteins were separated by SDS–PAGE (36) and transferred onto nitrocellulose membranes (Protran BA83; Schleicher & Schuell, Dassel, Germany) using a methanol-buffered solution (37). The transferred proteins were renatured at 4°C overnight in a hybridization buffer [0.01 M Tris–HCl, pH 7.5, 0.05 M NaCl, 1 mM EDTA, 2 mM dithiothreitol (DTT), 1× Denhardt's solution, 0.01% Tween-20] in the presence of 2 µg/ml *E.coli* tRNA (Sigma, St Louis, MO). The membranes were probed at room temperature for 1 h with radiolabelled RNA (100 000 c.p.m./ml) in a hybridization buffer containing 2–1000 µg/ml *E.coli* tRNA. The membranes were washed four times for 10 min in the hybridization buffer containing increasing NaCl concentrations and exposed to Kodak X-Omat AR films at –70°C using Dupont Lightning-Plus intensifying screens.

Ribonucleotide homopolymer-binding assays

Binding of the purified recombinant proteins to RNA homopolymers was performed as previously described (38). Briefly, each RNA homopolymer–agarose (100 µl; Sigma) was incubated with an equal amount of purified recombinant proteins, for 1 h at 4°C in a total volume of 1 ml of binding buffer (0.01 M Tris–HCl, pH 7.4, 0.05 M NaCl, 2.5 mM $MgCl_2$, 2 mM DTT). After incubation, the beads were split into three vials and washed five times for 10 min in 0.5 ml of ice-cold binding buffer containing 0.1, 0.25 or 0.5 M NaCl. Some washes were also performed in a binding buffer containing 1 mg/ml heparin and some binding reactions were supplemented with competitors. Proteins bound to the RNA homopolymer-bound beads were eluted by boiling in SDS loading buffer, resolved using SDS–PAGE and transferred to P-Immobilon membranes for immunoblotting with monoclonal anti-pentaHis antibody (Qiagen).

ABBREVIATIONS

Gem, gemini of coiled bodies; RBD, RNA-binding domain; RNP1, ribonucleoprotein consensus sequence motif 1; RNP2, ribonucleoprotein consensus sequence motif 2; SMA, spinal muscular atrophy; SMN, survival motor neuron; snRNP, small nuclear ribonucleoprotein.

ACKNOWLEDGEMENTS

We thank the patients, families and doctors without whom this study would not have been possible. We thank the organizers of the EMBO course on RNA–protein interactions 1998 and J. Marie for helpful discussions. We gratefully acknowledge H. Boulekbache (Université Paris 7-Diderot) for providing the zebrafish *Danio rerio* adult animals and Y. Deris for skilful

photographic assistance. We thank the members of the International SMA Consortium for stimulating discussions. S.B. is a recipient of an MRE Predoctoral Fellowship. This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM), Association Française contre les Myopathies (AFM), Action Concertée des Sciences du Vivant (ACC-SV2) and Programme Hospitalier de Recherche Clinique. Work in the laboratory of D.T.M. was supported by a grant from the Centre National de la Recherche Scientifique (CNRS).

REFERENCES

1. Munsat, T.L. (1991) Workshop report: international SMA collaboration. *Neuromusc. Disord*., **1**, 81.
2. Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Paslier, D., Frézal, J., Cohen, D., Weissenbach, J., Munnich, A. and Melki, J. (1995) Identification and characterization of a spinal muscular atrophy-determining gene. *Cell*, **80**, 155–165.
3. Burghes, A.H.M. (1997) When is a deletion not a deletion? When it is converted. Invited editorial. *Am. J. Hum. Genet*., **61**, 9–15.
4. Lefebvre, S., Bürglen, L., Frézal, J., Munnich, A. and Melki, J. (1998) The role of the *SMN* gene in proximal spinal muscular atrophy. *Hum. Mol. Genet*., **7**, 1531–1536.
5. Melki, J., Lefebvre, S., Bürglen, L., Burlet, P., Clermont, O., Millasseau, P., Reboullet, S., Bénichou, B., Zeviani, M., Le Paslier, D., Cohen, D., Weissenbach, J. and Munnich, A. (1994) *De novo* and inherited deletions of the 5q13 region in spinal muscular atrophy. *Science*, **264**, 1474–1477.
6. Viollet, L., Bertrandy, S., Brunialti, A.L.B., Lefebvre, S., Burlet, P., Clermont, O., Cruaud, C., Guénet, J.L., Munnich, A. and Melki, J. (1997) cDNA isolation, expression and chromosomal localization of the mouse survival motor neuron gene (*Smn*). *Genomics*, **40**, 185–188.
7. DiDonato, C.J., Chen, X.N., Kerenberg, J.R., Nadeau, J.H. and Simard, L.R. (1997) Cloning, characterization and copy number of the murine survival motor neuron gene, homolog of the spinal muscular atrophy-determining gene. *Genome Res*., **7**, 339–352.
8. Bürglen, L., Lefebvre, S., Clermont, O., Burlet, P., Viollet, L., Cruaud, C., Munnich, A. and Melki, J. (1996) Structure and organization of the human survival motor neuron (SMN) gene. *Genomics*, **32**, 479–482.
9. Coovert, D.D., Le, T.T., McAndrew, P.E., Strasswimmer, J., Crawford, T.O., Mendell, J.R., Coulson, S., Androphy, E.J., Prior, T.W. and Burghes, A.H.M. (1997) The survival motor neuron protein in spinal muscular atrophy. *Hum. Mol. Genet*., **6**, 1205–1214.
10. Lefebvre, S., Burlet, P., Liu, Q., Bertrandy, S., Clermont, O., Munnich, A., Dreyfuss, G. and Melki, J. (1997) Correlation between severity and SMN protein level in spinal muscular atrophy. *Nature Genet*., **16**, 265–269.
11. Burlet, P., Huber, C., Bertrandy, S., Ludosky, M.A., Zwaenepoel, I., Clermont, O., Roume, J., Delezoide, A.L., Cartaud, J., Munnich, A. and Lefebvre, S. (1998) The distribution of SMN protein complex in human fetal tissues and its alteration in spinal muscular atrophy. *Hum. Mol. Genet*., **7**, 1927–1933.
12. Liu, Q. and Dreyfuss, G. (1996) A novel nuclear structure containing the survival of motor neurons protein. *EMBO J*., **15**, 3555–3565.
13. Liu, Q., Fischer, U., Wang, F. and Dreyfuss, G. (1997) The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell*, **90**, 1013–1021.
14. Fischer, U., Liu, Q. and Dreyfuss, G. (1997) The SMN–SIP1 complex has an essential role in the spliceosomal snRNP biogenesis. *Cell*, **90**, 1023–1029.
15. Pellizoni, L., Kataoka, N., Charroux, B. and Dreyfuss, G. (1998) A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell*, **95**, 615–624.
16. Talbot, K., Ponting, C.P., Theodosiou, A.M., Rodrigues, N.R., Surtees, R., Mountford, R. and Davies, K.E. (1997) Missense mutation clustering in the survival motor neuron gene. A role for a conserved tyrosine and glycine rich region of the protein in RNA metabolism? *Hum. Mol. Genet*., **6**, 497–501.
17. Lorson, C.L., Strasswimmer, J., Yao, J.M., Baleja, J.D., Hahnen, E., Wirth, B., Thanh, L., Burghes, A.H.M. and Androphy, E.J. (1998) SMN oligomerization defect correlates with spinal muscular atrophy. *Nature Genet*., **19**, 63–66.
18. Lorson, C.L. and Androphy, E.J. (1998) The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. *Hum. Mol. Genet*., **7**, 1269–1275.
19. Burd, C. and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.
20. Landsman, D. (1992) RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. *Nucleic Acids Res*., **20**, 2861–2864.
21. Tabara, H., Motohashi, T. and Kohara, Y. (1996) A multi-well version of *in situ* hybridization on whole mount embryos of *Caenorhabditis elegans*. *Nucleic Acids Res*., **24**, 2119–2124.
22. Blumenthal, T. (1995) Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet*., **11**, 132–136.
23. Gennarelli, M., Lucarelli, M., Capon, F., Pizzuti, A., Merlini, L., Angelini, C., Novelli, G. and Dallapiccola, B. (1995) Survival motor neuron gene transcript analysis in muscles from spinal muscular atrophy patients. *Biochem. Biophys. Res. Commun*., **213**, 342–348.
24. Ponting, C. (1997) Tudor domains in proteins that interact with RNA. *Trends Biochem. Sci*., **22**, 51–52.
25. Brahe, C., Clermont, O., Zappata, S., Tiziano, F., Melki, J. and Neri, G. (1996) Frameshift mutation in the survival motor neuron gene in a severe case of SMA type I. *Hum. Mol. Genet*., **5**, 1971–1976.
26. Schrank, B., Götz, R., Gunnersen, J.M., Ure, J.M., Toyka, K.V., Smith, A.G. and Sendtner, M. (1997) Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos. *Proc. Natl Acad. Sci. USA*, **94**, 9920–9925.
27. The *C.elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C.elegans*: a platform for investigating biology. *Science*, **282**, 2012–2018.
28. Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh P., Sleeman, J., Lamond, A. and Mann, M. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nature Genet*., **20**, 46–50.
29. Talbot, K., Miguel-Aliaga, I., Mohaghegh, P., Ponting, C.P. and Davies, K.E. (1998) Characterization of a gene encoding survival motor neuron (SMN)-related protein, a constituent of the spliceosome complex. *Hum. Mol. Genet*., **7**, 2149–2156.
30. Varani, G. and Nagai, K. (1998) RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct*., **27**, 407–445.
31. Mattaj, I.W. (1998) Ribonucleoprotein assembly: clues from spinal muscular atrophy. *Curr. Biol*., **8**, R93–R95.
32. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*., **22**, 4673–4680.
33. Bairoch, A. (1993) The PROSITE dictionary of sites and patterns in proteins, its current status. *Nucleic Acids Res*., **21**, 3097–3103.
34. Sonnhammer, E.L. and Kahn, D. (1994) Modular arrangement of proteins as inferred from analysis of homology. *Protein Sci*., **3**, 482–492.
35. Fondrat, C. and Dessen, P. (1995) A rapid access motif database (RAMdb) with a search algorithm for the retrieval patterns in nucleic acids or protein databanks. *CABIOS*, **11**, 273–279.
36. Laemmli, U.K. (1970) Cleavage of structural proteins during assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
37. Towbin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl Acad. Sci. USA*, **76**, 4350–4354.
38. Swanson, M.S. and Dreyfuss, G. (1988) Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol. Cell. Biol*., **8**, 2237–2241.